



Seeing the forest, not the trees: what can we learn from the study of in vivo animal research?

Malcolm Macleod

Centre for Clinical Brain Sciences, University of Edinburgh

slides at @CAMARADES_



Why do we do in vivo experiments?

Understand

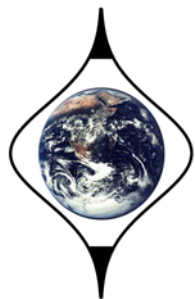
- Understand what causes the disease
- Understand which biological processes are pivotal and which are not

Influence in models

- Be able to change these processes in experiments
- Be able to change outcome in disease models
- Know that our treatment is probably safe

Prevent in real life

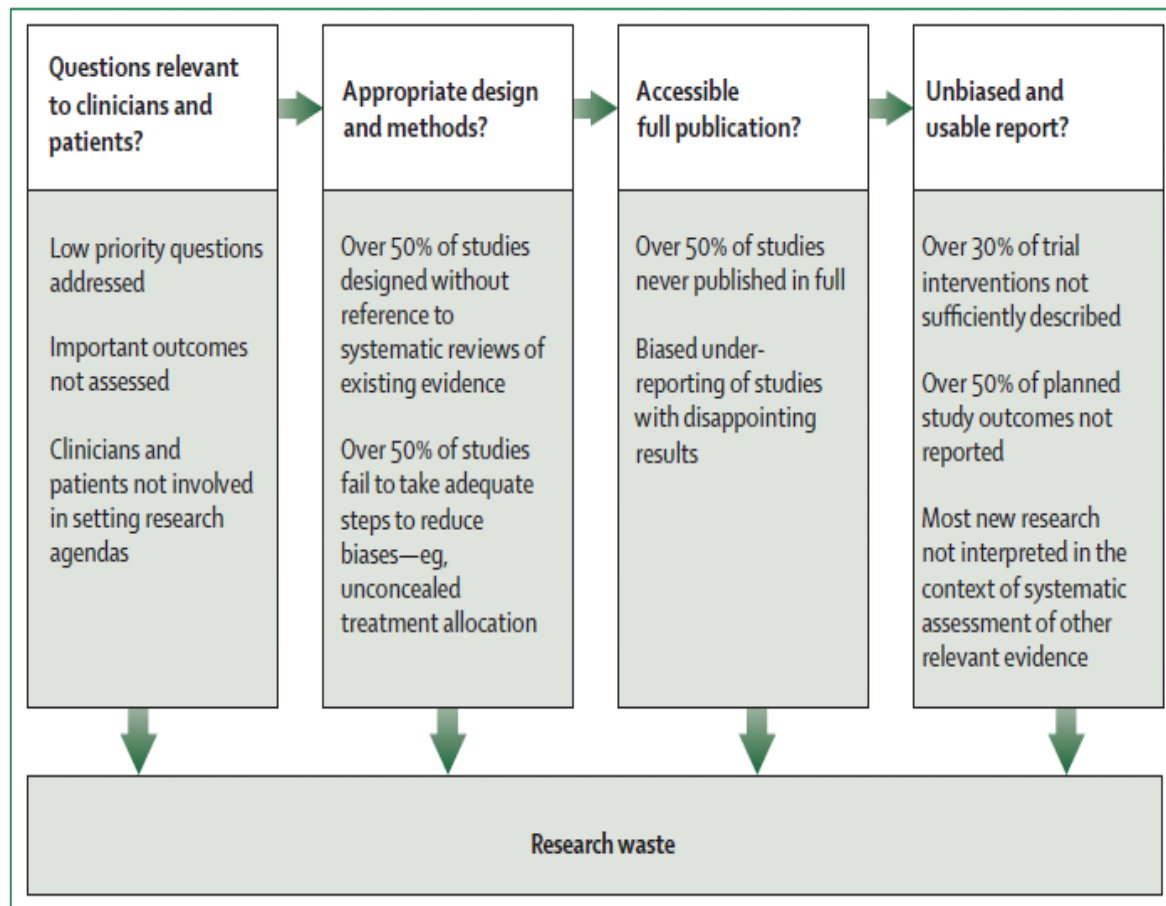
- Show, in clinical trials, that the treatment changes outcome
- Show that the treatment works in the real world



Avoidable waste in the production and reporting of research evidence



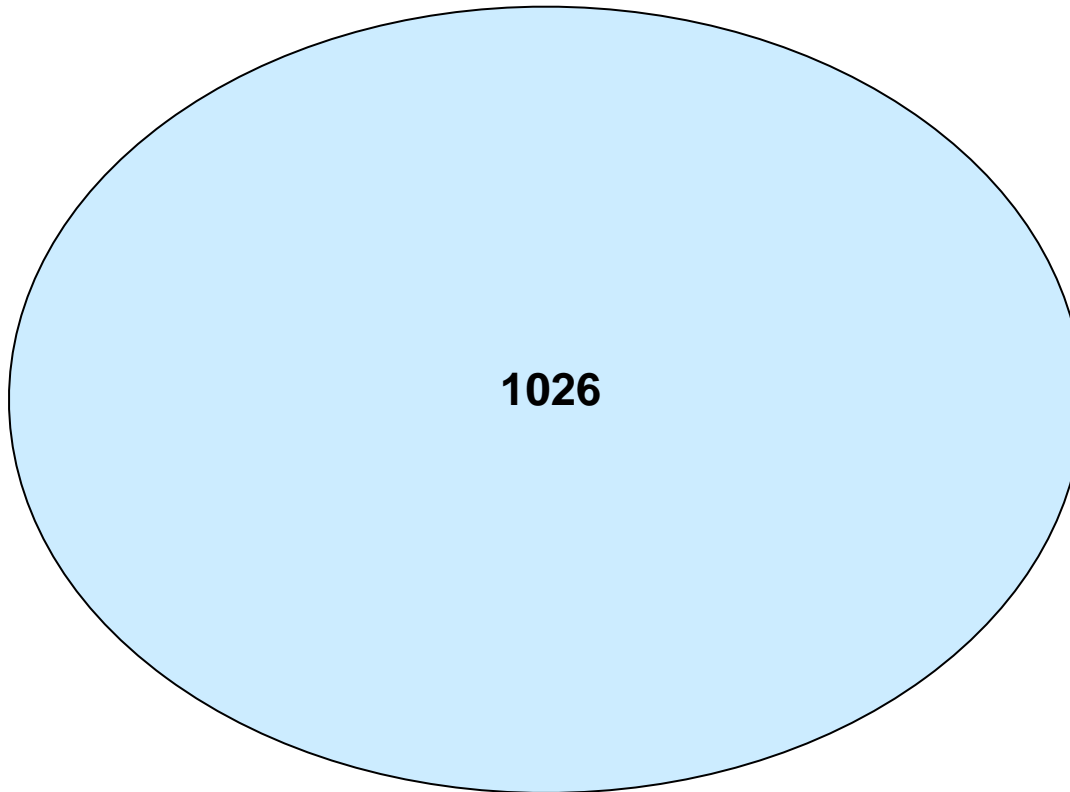
Iain Chalmers, Paul Glasziou



Lancet 2009; 374: 86–89



1026 interventions in experimental stroke



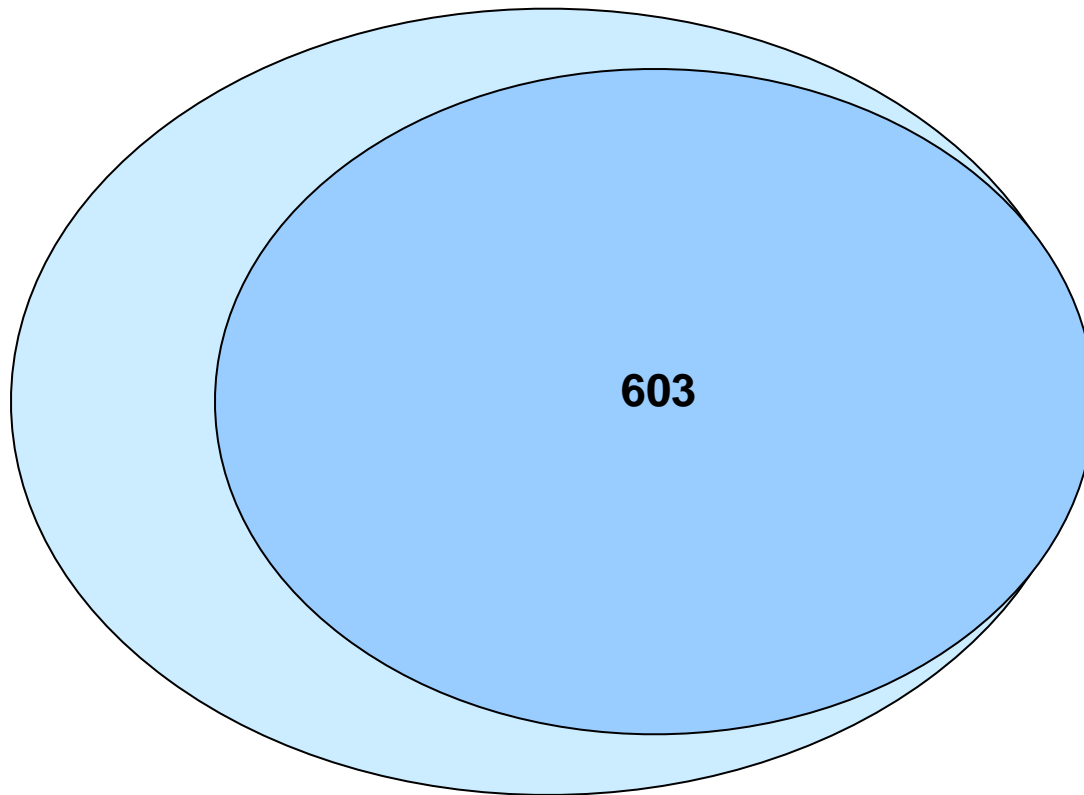
Developed in *in vitro* and *in vivo* experiments

O' Collins et al, 2006

CAMARADES: Bringing evidence to translational medicine



1026 interventions in experimental stroke

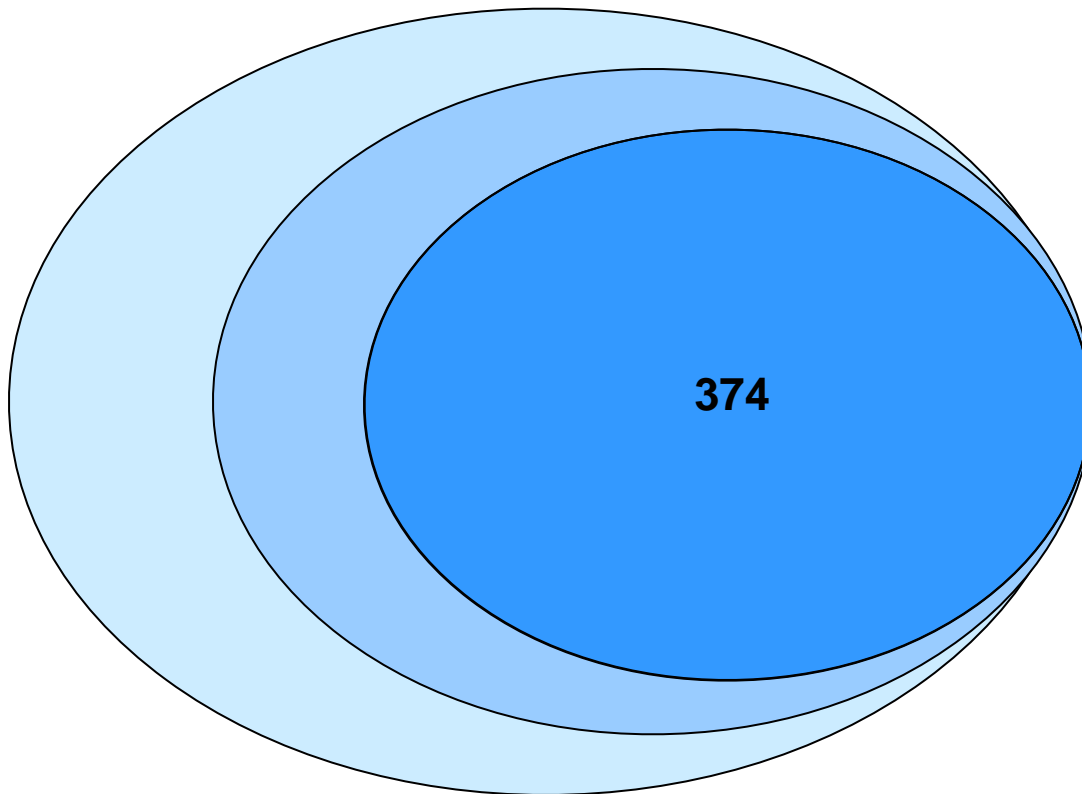


Tested in focal ischaemia

O' Collins et al, 2006



1026 interventions in experimental stroke

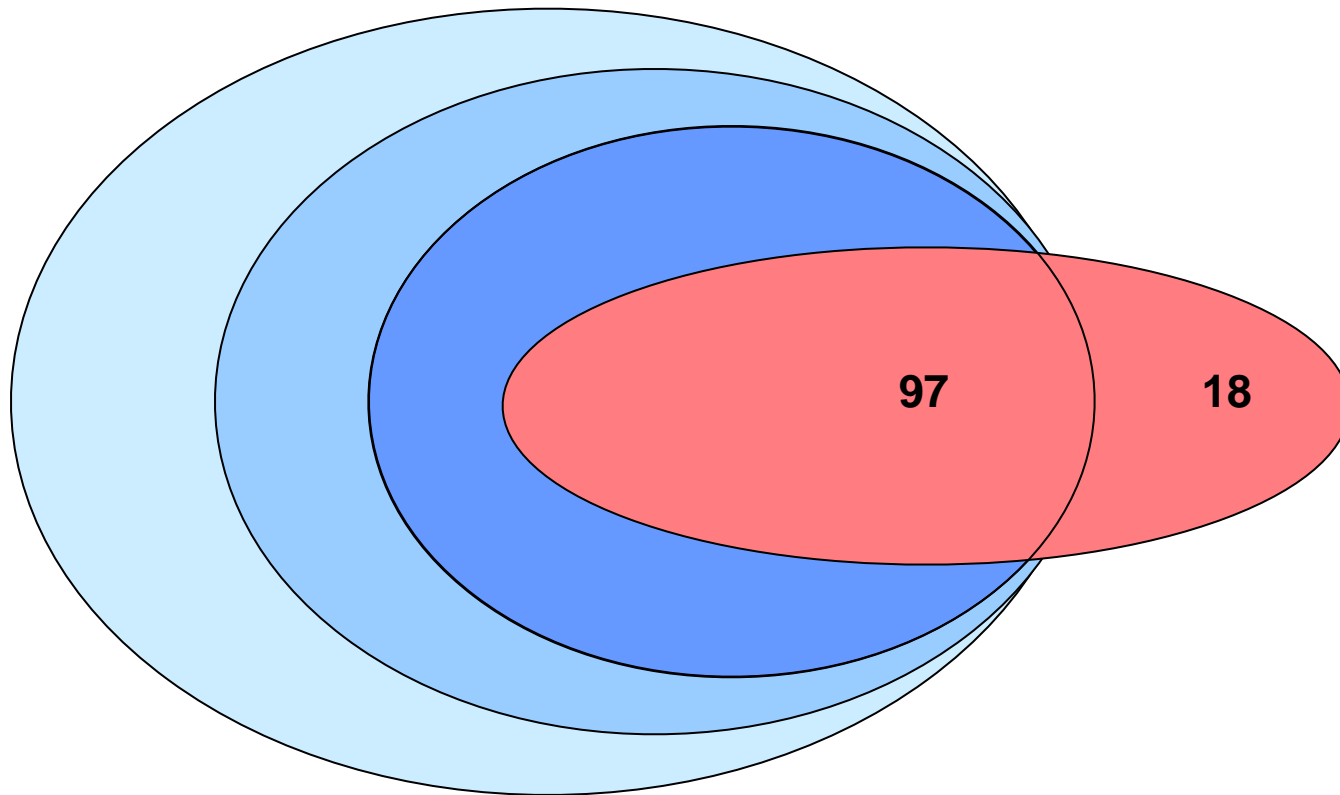


Effective in focal ischaemia

O' Collins et al, 2006



1026 interventions in experimental stroke

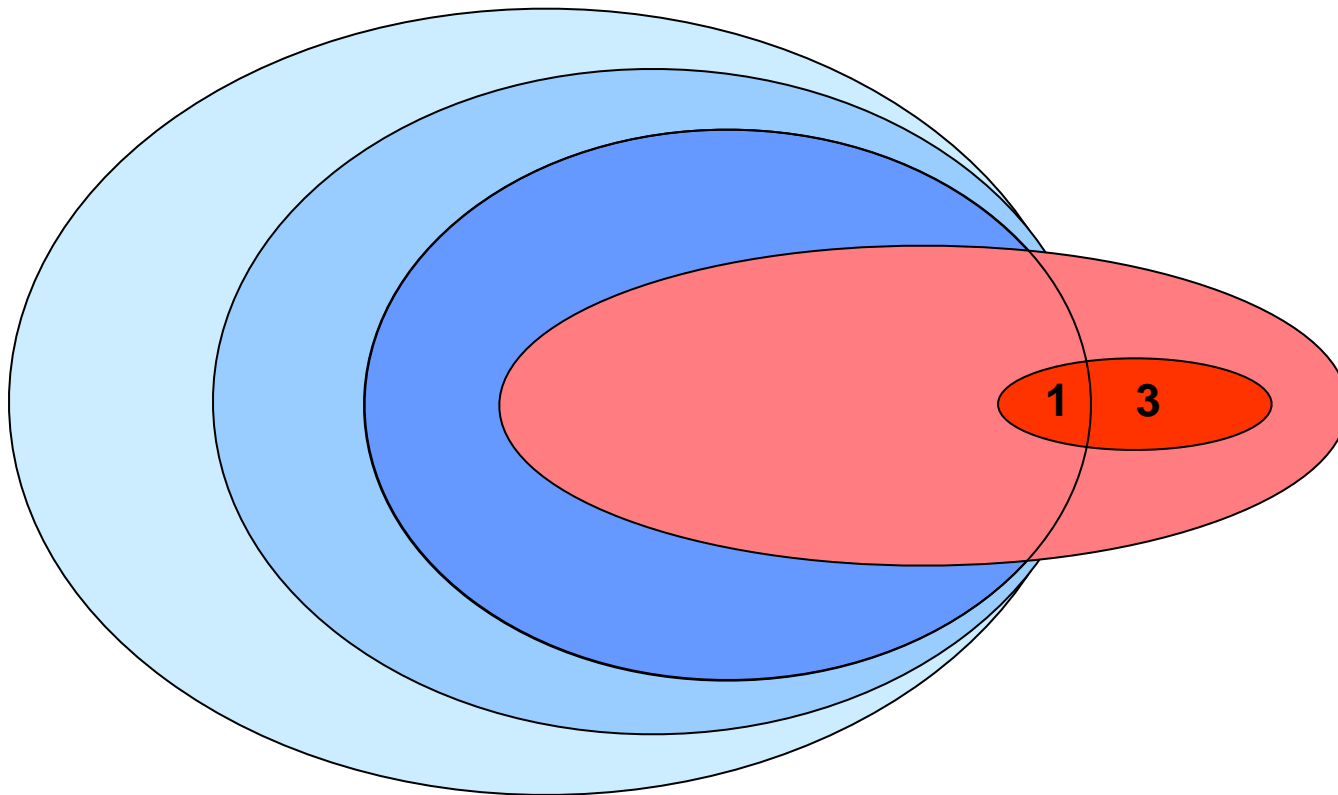


Tested in clinical trial

O' Collins et al, 2006

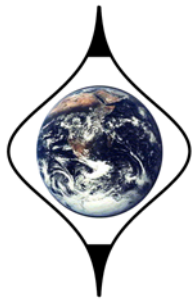


1026 interventions in experimental stroke



Effective in clinical trial

O' Collins et al, 2006



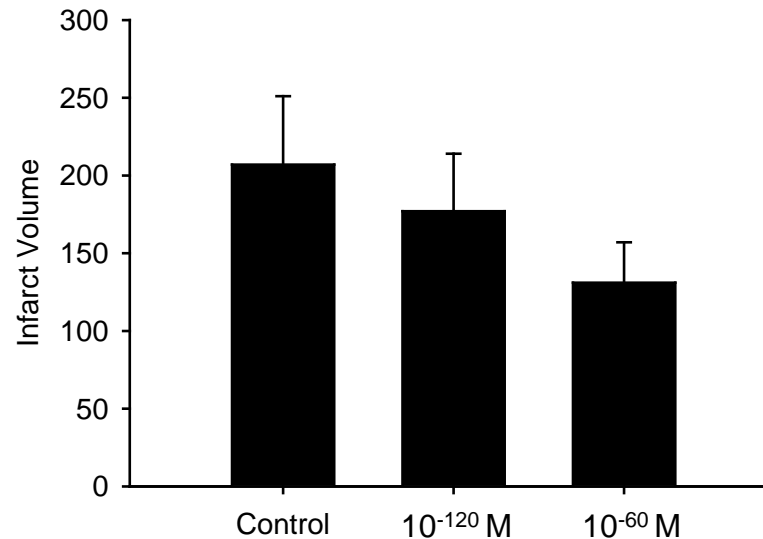
Treatment of experimental stroke with low-dose glutamate and homeopathic *Arnica montana**

W. Jonas¹, Y. Lin², A. Williams², F. Tortella², R. Tuma³

¹ Uniformed Services University of the Health Sciences, Bethesda, Maryland

² Walter Reed Army Institute of Research, Washington, D.C.

³ Temple University, Philadelphia, PA

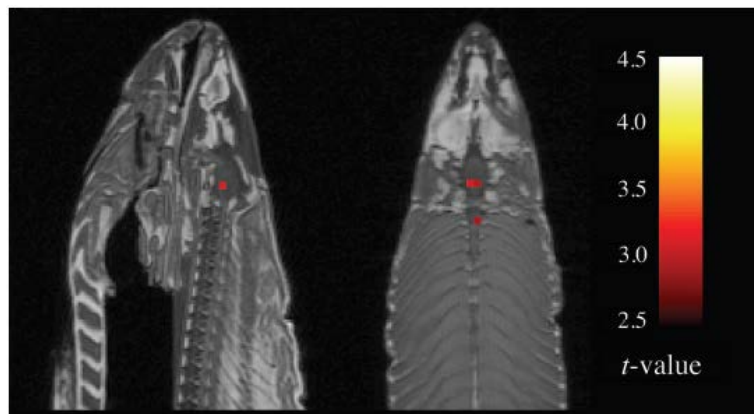




Neural Correlates of Interspecies Perspective Taking in the Post-Mortem Atlantic Salmon: An Argument For Proper Multiple Comparisons Correction

Craig M. Bennett^{1*}, Abigail A. Baird², Michael B. Miller¹ and George L. Wolford³

One mature Atlantic Salmon (*Salmo salar*) participated in the fMRI study. The salmon measured approximately 18 inches long, weighed 3.8 lbs, and was not alive at the time of scanning. It is not known if the salmon was male or female, but given the post-mortem state of the subject this was not thought to be a critical variable.



The task administered to the salmon involved completing an open-ended mentalizing task. The salmon was shown a series of photographs depicting human individuals in social situations with a specified emotional valence, either socially inclusive or socially exclusive. The salmon was asked to determine which emotion the individual in the photo must have been experiencing.

Several active voxels were observed in a cluster located within the salmon's brain cavity (see Fig. 1). The size of this cluster was 81 mm³ with a cluster-level significance of $p = 0.001$.

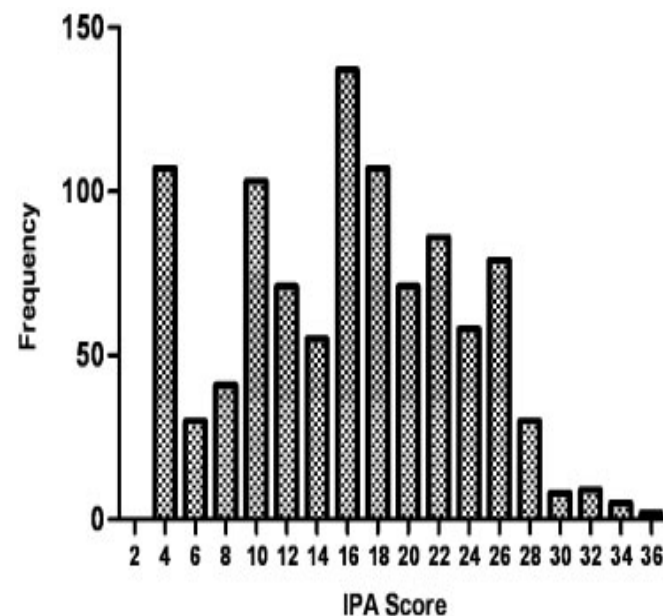
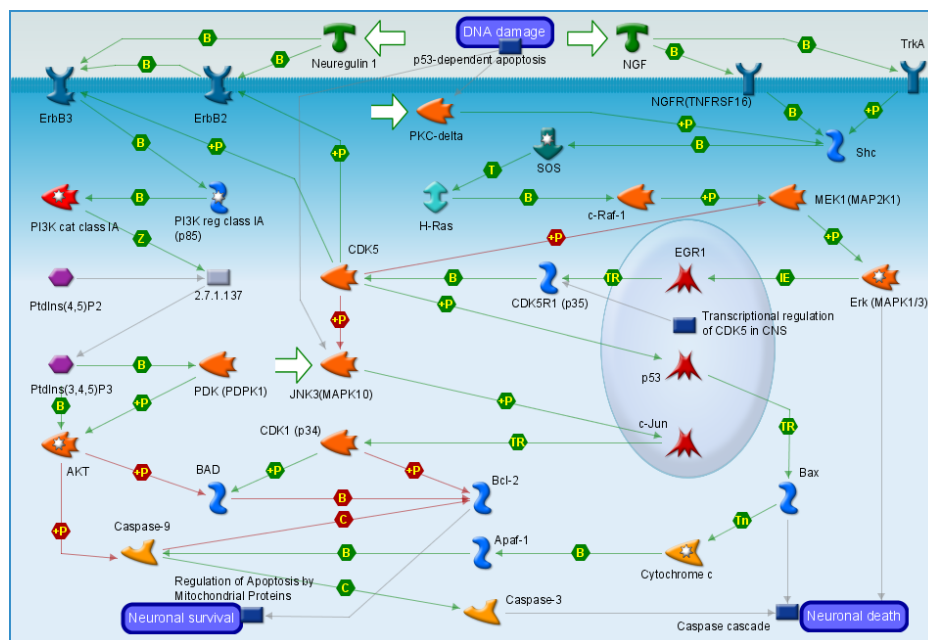
Either we have stumbled onto a rather amazing discovery in terms of post-mortem ichthyological cognition, or there is something a bit off with regard to our uncorrected statistical approach.



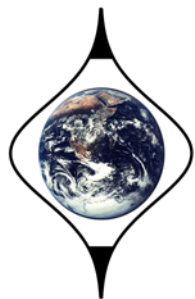
Winner of the 2012 Ignoble Prize for Neuroscience



Connecting chains of evidence



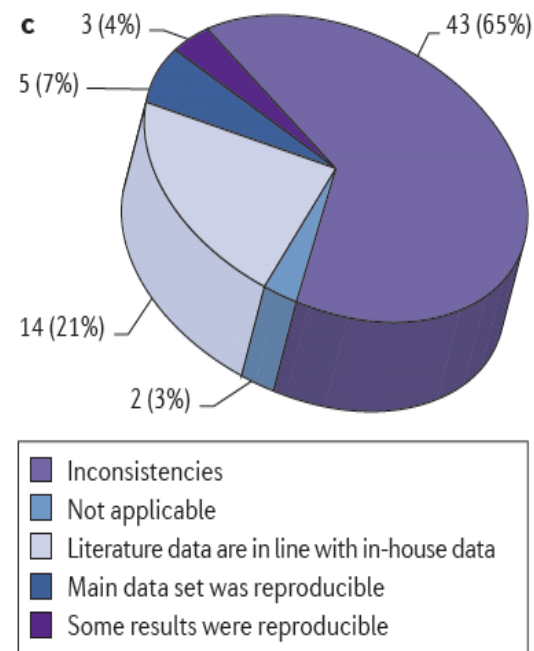
Deighton et al 2010 Figure 2: Bootstrap of 1000 sets of 13 random proteins. A frequency distribution of the IPA scores from 1000 randomly generated sets of 13 proteins. The median IPA score is 16. Only 16 of the 1000 random sets have a score of 30 or greater.



What happens when pharma tries to replicate academic findings?



- Bayer, Berlin
- In-house target identification and validation projects over 4 years
- Womens' health, cardiovascular disease, oncology
- 67 projects



Prinz et al, Nature Reviews Drug Discovery, 2011



Potential sources of bias in animal studies



Internal Validity	Solution
Selection Bias	Randomisation
Performance Bias	Allocation Concealment
Detection Bias	Blinded outcome assessment
Attrition bias	Reporting drop-outs/ ITT analysis
External Validity	
Publication bias	Registries, protocols
Relevance of models	Appropriate times, co-morbidities

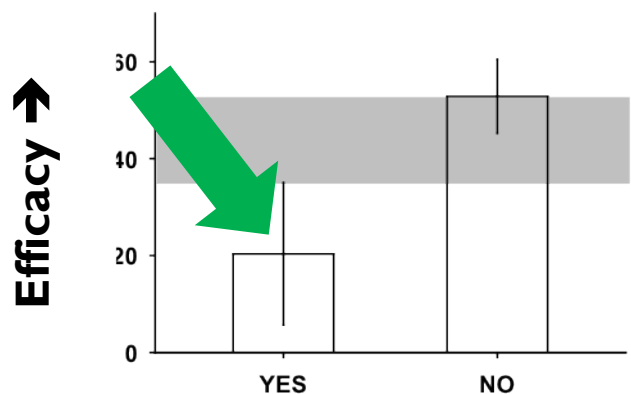


Why should translational medicine be evidence-based?

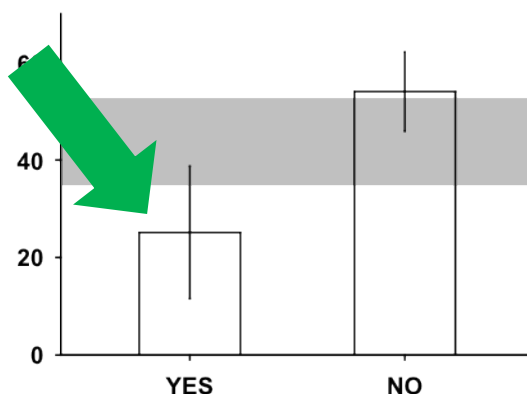
Lessons from NXY-059



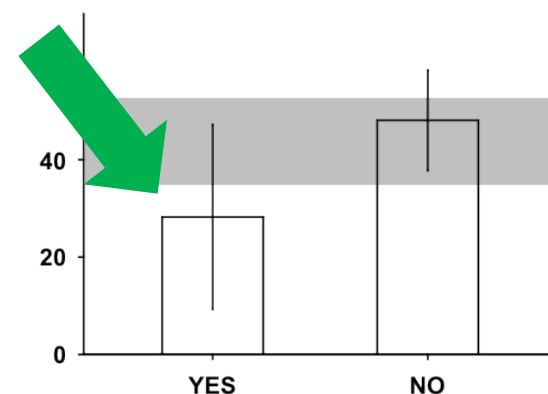
- Infarct Volume
 - 11 publications, 29 experiments, 408 animals
 - Improved outcome by 44% (35-53%)



Randomisation



**Blinded conduct
of experiment**



**Blinded
assessment of
outcome**

Macleod et al, 2008

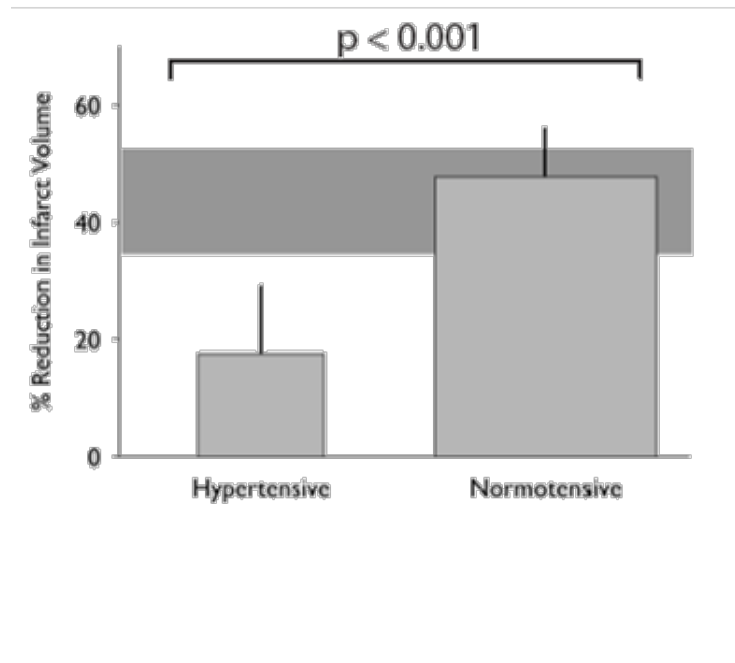


External validity in stroke modelling - Hypertension and NXY-059



Hypertension:

- 7% of animal studies
- 77% of patients in the (neutral) SAINT II study

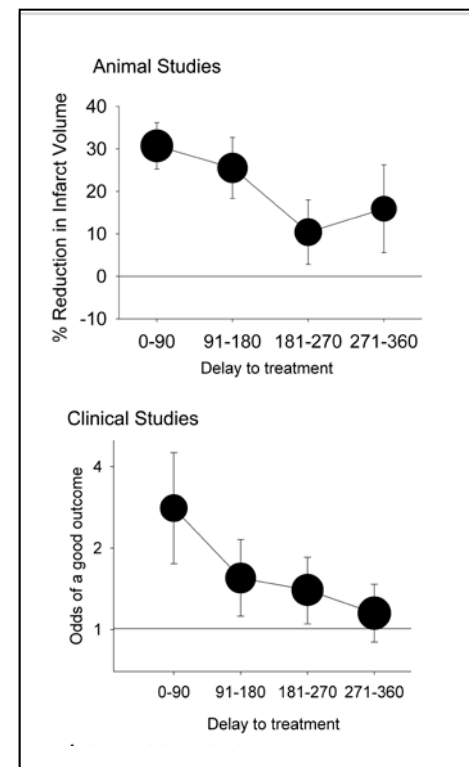




Animal studies can be concordant with clinical trials



- **Both tPA and tirilazad appear to work in animals**
- **tPA works in humans but tirilazad doesn't**
- **Time to treatment: tPA:**
 - Animals – median 90 minutes
 - Clinical trial – median 90 minutes
- **Time to treatment: tirilazad**
 - Animals – median 10 minutes
 - Clinical trial – >3 hrs for >75% of patients

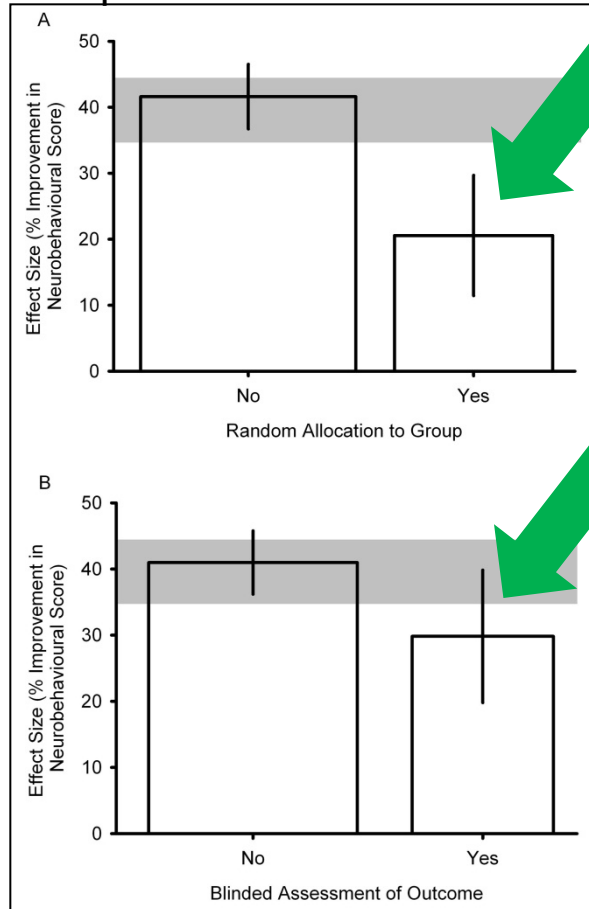




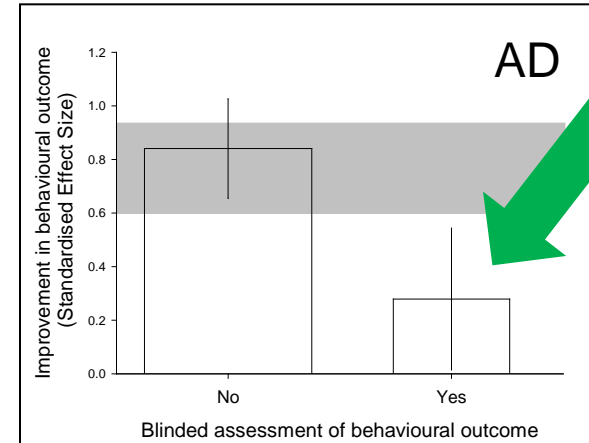
Lessons from other neuroscience domains



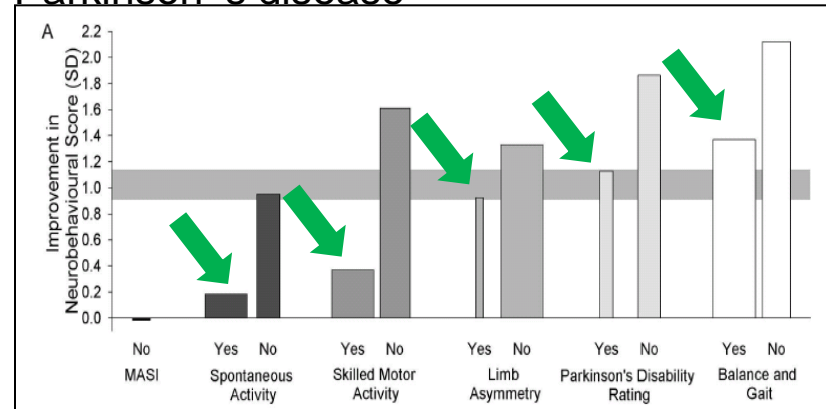
Multiple Sclerosis



Alzheimer's disease



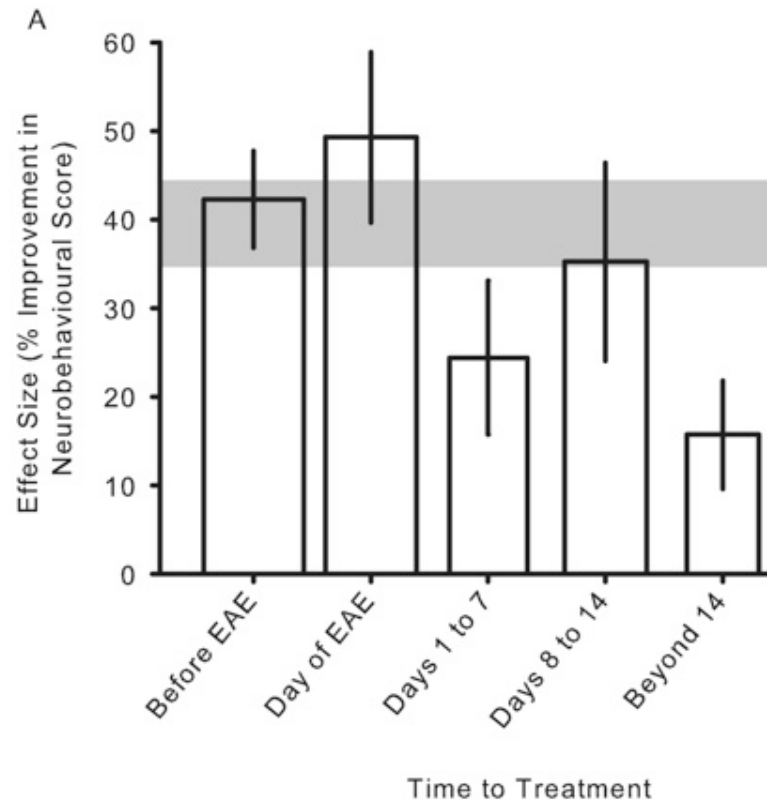
Parkinson's disease





External validity

Timing of treatment





Publication bias



20%

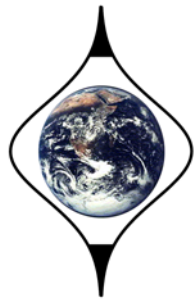
- 32%

	n expts	Estimated unpublished	Reported efficacy	Corrected efficacy
Stroke – infarct volume	1359	214	31.3%	23.8%
EAE - neurobehaviour	1892	505	33.1%	15.0%
EAE – inflammation	818	14	38.2%	37.5%
EAE – demyelination	290	74	45.1%	30.5%
EAE – axon loss	170	46	54.8%	41.7%
AD – Water Maze	80	15	0.688 sd	0.498 sd
AD – plaque burden	632	154	0.999 sd	0.610 sd



Excess significance

- 4,445 study datasets synthesized in 160 meta-analyses
 - Alzheimer's disease, EAE, focal ischemia, intracerebral hemorrhage, Parkinson's disease, and spinal cord injury.
- Expected significant results = 919 (21%)
- Observed significant results = 1719 (39%)
- Excess significance was present across all neurological disorders, in all subgroups defined by methodological or reporting characteristics.
- Observed effective interventions in 112 (70%) of meta-analyses
- Significantly effective interventions with more than 500 animals and no hints of bias were seen only in eight (5%) of the 160 meta-analyses.



Reporting of measures to avoid bias



- Kilkenney et al (2009)
 - 271 papers published 1999-2005
 - “publically funded research”
 - Sample size calculation 0/48 (0% of those examined in detail)
 - Randomisation 32/271 (12%)
 - Blinding 5/35 (14% of those reporting qualitative data)
- Vesterinen et al (2011)
 - All original research in 2008 volume of JCBFM
 - Sample size calculation 2/311 (1%)
 - Randomisation 46/292 (16%)
 - Blinding 46/312 (15%)



A year in the life of JCBFM

Reporting of measures to avoid bias



Item	<i>In vitro</i> (n=98)	Animal (n=190)	Human (n=68)	Total (n = 312)
Randomisation	15%	22%	8%	16%
Allocation Concealment	2%	8%	5%	6%
Blinded assessment of outcome	13%	15%	25%	15%
Sample size calculation	2%	1%	0%	1%



What have we learned from systematic reviews of interventions?



- There is a low prevalence of reporting of measures which might increase validity
- Studies not reporting such measures give higher estimates of efficacy
- Clinical trials may have been based on too optimistic an assessment of preclinical efficacy



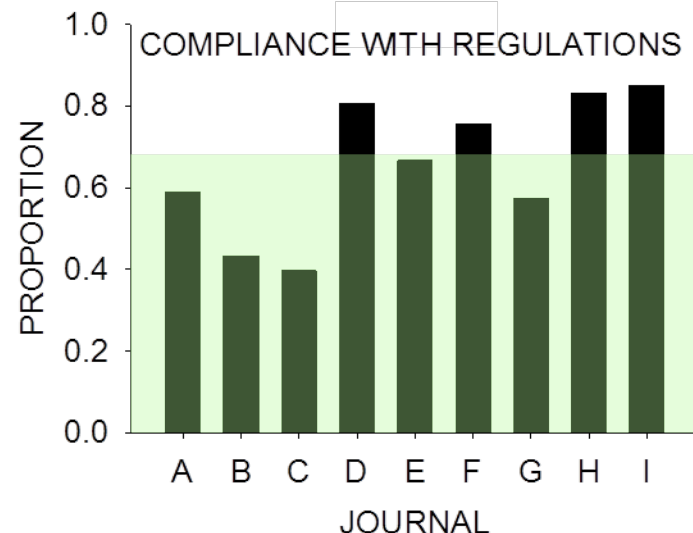
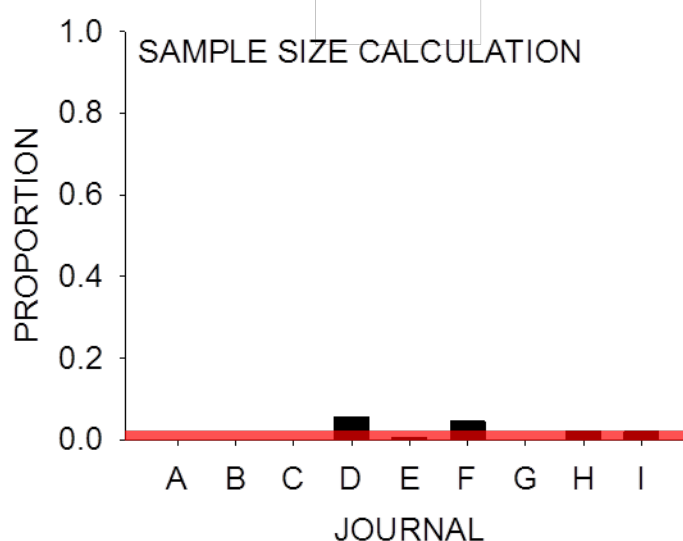
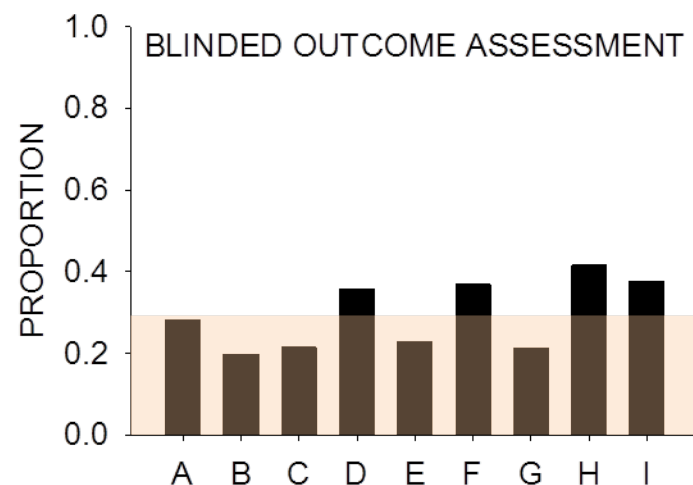
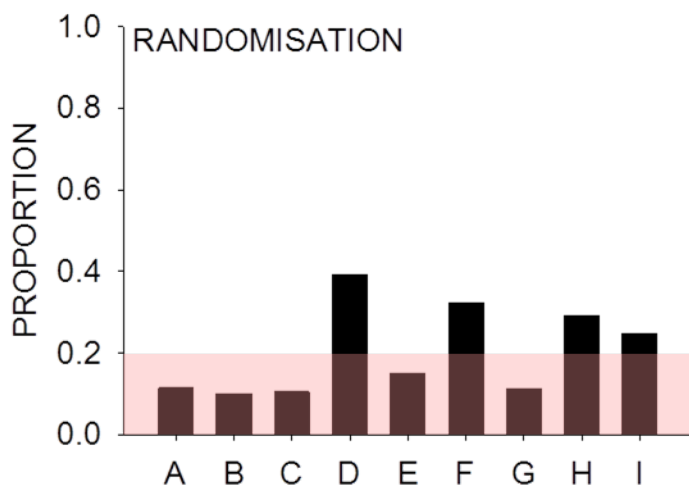
Prevalence of some measures to improve validity



	Randomisation	Blinded Outcome Assessment	Sample Size calculation
Stroke	36%	29%	3%
MND	31%	20%	<1%
AD	15%	25%	0%
PD	12%	15%	0%
EAE	8%	15%	<1%
Glioma	14%	0%	0%
Pain	14%	25%	0%



4 different quality items ...

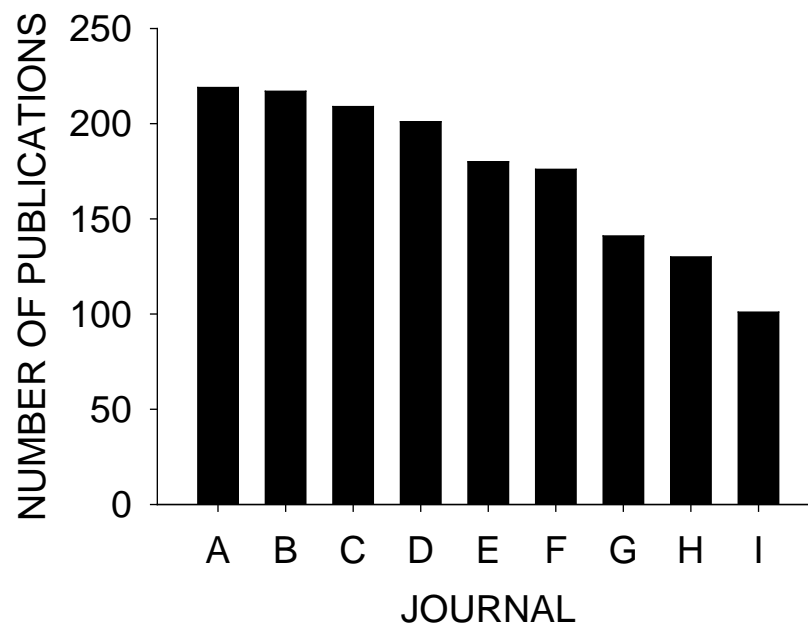




Quality is different in different Journals

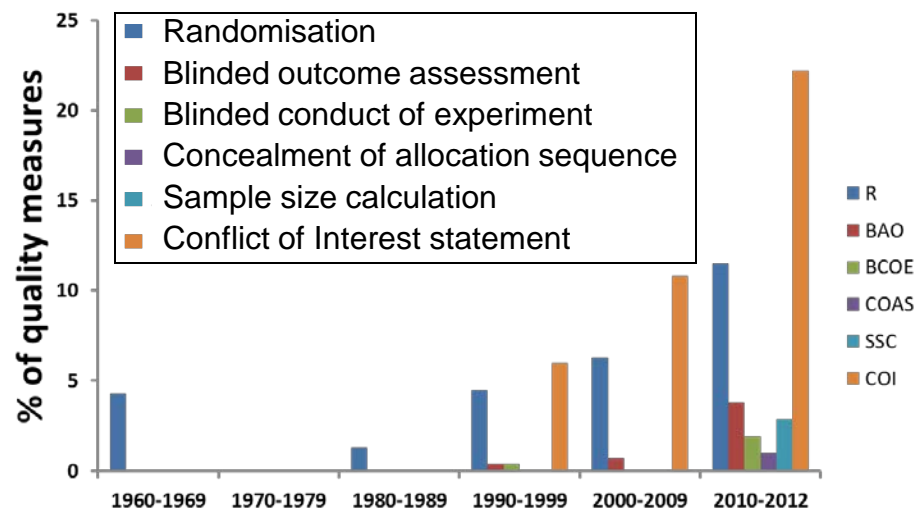


- 4584 full publications curated on CAMARADES
- Studies reporting the efficacy of an intervention in an animal disease model
- Limited to journals contributing more than 100 publications
 - Brain Research
 - Experimental Neurology
 - JCBFM*
 - Journal of Immunology
 - Journal of Neuroimmunology
 - Journal of Neuroscience
 - Neuroscience
 - PNAS
 - Stroke



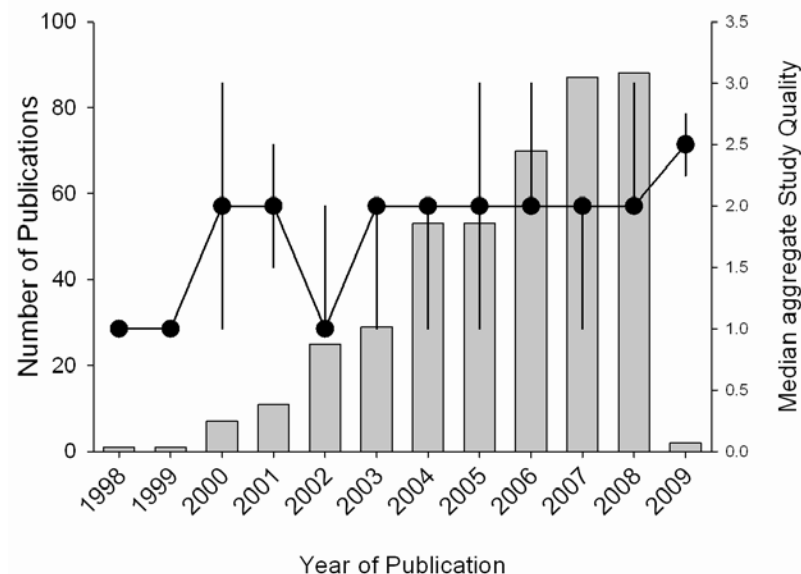


Has quality improved over time?



Study quality of *in vivo* studies selected from random sample of 1000 publications from PubMed

- EAE: some improvement over time: 26 years per point increment in quality
- AD: some improvement over time: 24 years per point improvement in study quality



CUMULATIVE META-ANALYSIS OF THERAPEUTIC TRIALS FOR MYOCARDIAL INFARCTION

JOSEPH LAU, M.D., ELLIOTT M. ANTMAN, M.D., JEANETTE JIMENEZ-SILVA, M.D., BRUCE KUPELNICK, B.A.,
FREDERICK MOSTELLER, PH.D., AND THOMAS C. CHALMERS, M.D.

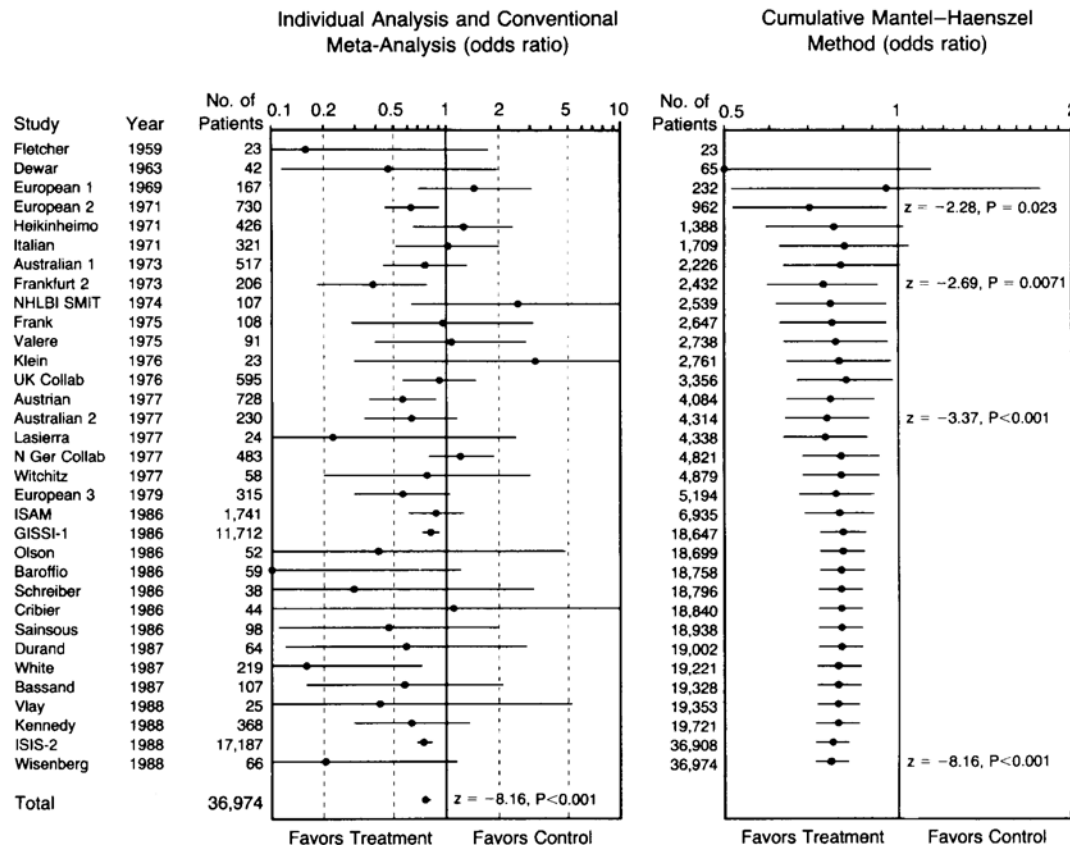


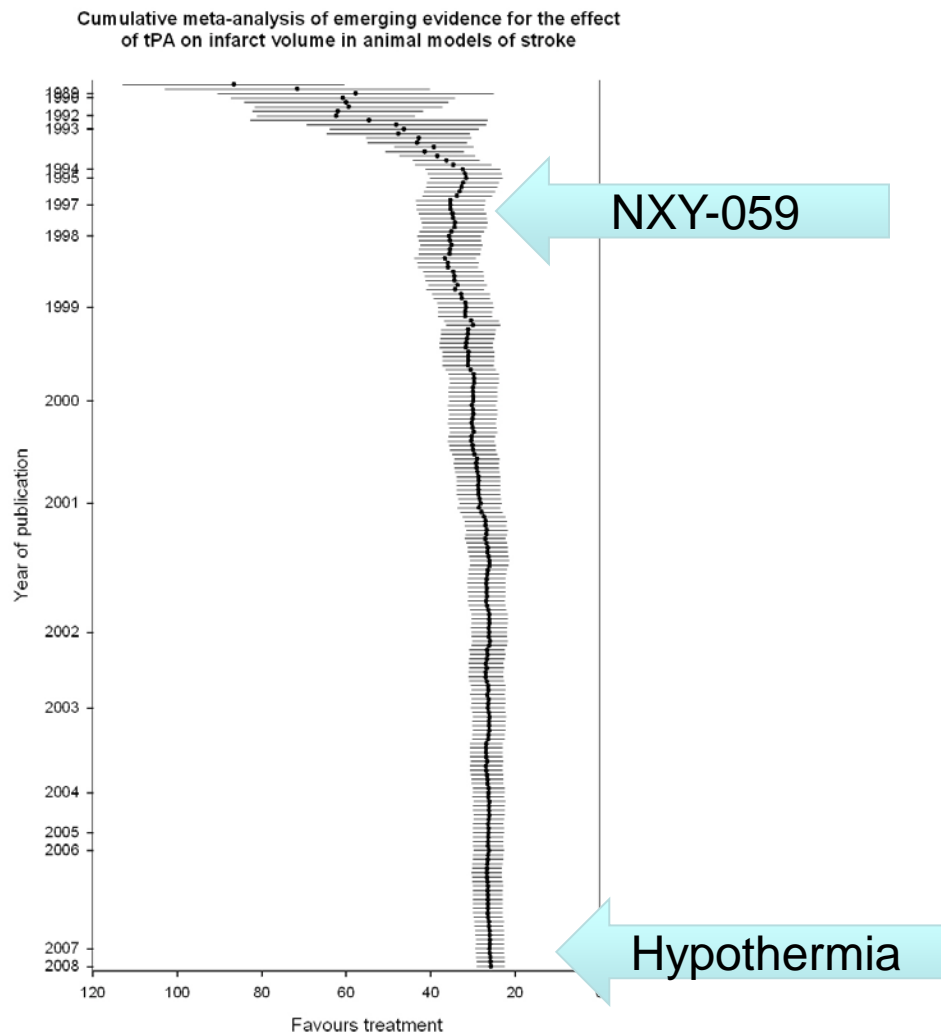
Figure 1. Conventional and Cumulative Meta-Analyses of 33 Trials of Intravenous Streptokinase for Acute Myocardial Infarction. The odds ratios and 95 percent confidence intervals for an effect of treatment on mortality are shown on a logarithmic scale. A bibliography of the published trial reports is available from the authors.



Data: More are better



Cumulative meta-analysis of the efficacy of lytic treatments (eg tPA) in thrombotic animal models of stroke





So what's new?



“We must be just as diligent in seeking data contrary to our hypothesis as we are in ferreting out data that may support it. Let us avoid excessive attachment to our own ideas, which we need to treat as prosecutor, not defense attorney. Even though a tumor is ours, it must be removed. It is far better to correct ourselves rather than to endure correction by others”

Santiago Ramon y Cahal (1898)
“Advice to the young investigator”



“The records of two investigators will not dovetail exactly, even when they read figures from a dial. Errors may creep in, and the direction of the error is more likely than not to be associated with the observer’s interest in how the findings come out”

Anne Roe, *The Psychology of the Scientist*,
Science, 134, 456-9, **1961**

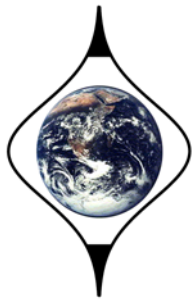


Expectancy effects in rats

- 12 students enrolled in laboratory course in experimental psychology
- Students matched for how much they thought they would like working with rats
- Rats matched on age
- Rats in T maze with dark arm alternating at random and the dark arm always reinforced
- Ten trials a day for 5 days
- Number of correct responses recorded

Group	Day 1	Day 2	Day 3	Day 4	Day 5	Mean
"Maze bright"	1.33	1.60	2.60	2.83	3.26	2.32
"Maze dull"	0.72	1.10	2.23	1.83	1.83	1.54
Δ	+0.60	+0.50	+0.37	+1.00	+1.43	+0.78
t	2.54	1.02	0.29	2.28	2.37	4.10
p	.03	.18	.39	.04	.03	.01

Rosenthal and Fode, Behav Sci 8, 183-9



...experimenter expectancies may be significant determinants of the results of their research employing animal subjects (experiments) suggest that the mediation of this expectancy biasing phenomenon may be extremely subtle”

Rosenthal (**1961**), Experimenter Effects in Behavioral Research



Science is not a profession ...



	Specialist training	Regulatory Body	Code of Professional Ethics	Continuing Professional Development
Science	✓	X	X	X
Medicine	✓	✓	✓	✓
Law	✓	✓	✓	✓
Teaching	✓	✓	✓	✓
Nursing	✓	✓	✓	✓
Football	✓	✓	✓	✓



What should scientists do?

- Be rigorous in demanding real sample size calculations
- Be rigorous in demanding the highest quality standards in the conduct and reporting of studies
- Develop model specific guidelines for good laboratory practice
- Develop registries of animal studies to prevent unnecessary replication and to address publication bias
- Where effect sizes are small or where human trial planned on basis of animal data, develop tools for multicentre animal studies
- Develop codes of professional practice
- Develop supportive, beaurocracy-lite capability for CPD, appraisal, validation and revalidation



Implications for toxicology studies



- In preclinical experiments
 - The bias – of investigators and funders – is likely to be towards overstatement of effect sizes
 - Publication bias, where it is present, is likely to lead to suppression of neutral or negative studies
- In toxicology studies
 - The bias – of funders and potentially of investigators – is likely to be towards understatement of effect sizes
 - Publication bias – where it is present – may lead to a suppression of positive studies (q.v. adverse effects in clinical trials)